Multi-Petabyte Data Distribution in Industry & Science with CernVM File System
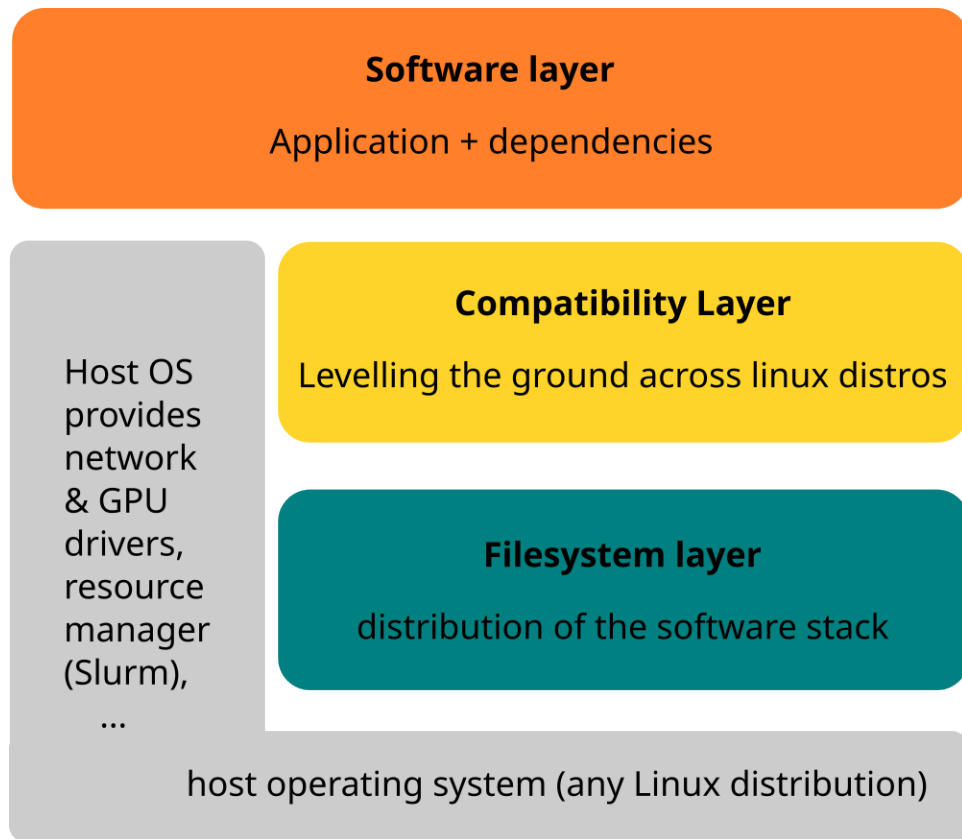
Georgios Christodoulis, CERN
Andriy Utkin

CernVM
File system

1

Software distribution in Particle Physics

European Environment for Scientific Software Installations (EESSI)

CernVM Filesystem (CVMFS)

```
[root@christge-cl ~]# tree -L 2 /cvmfs/atlas.cern.ch/
/cvmfs/atlas.cern.ch/
├── repo
│   ├── ATLASLocalRootBase
│   ├── benchmarks
│   ├── conditions -> /cvmfs/atlas-condb.cern.ch/repo/conditions
│   ├── containers
│   ├── dev
│   ├── sw
│   ├── test
│   ├── tools
│   └── tutorials
└── test1

10 directories, 2 files
[root@christge-cl ~]# cat /cvmfs/atlas.cern.ch/test1 #JIT download test1
test
[root@christge-cl ~]#
```
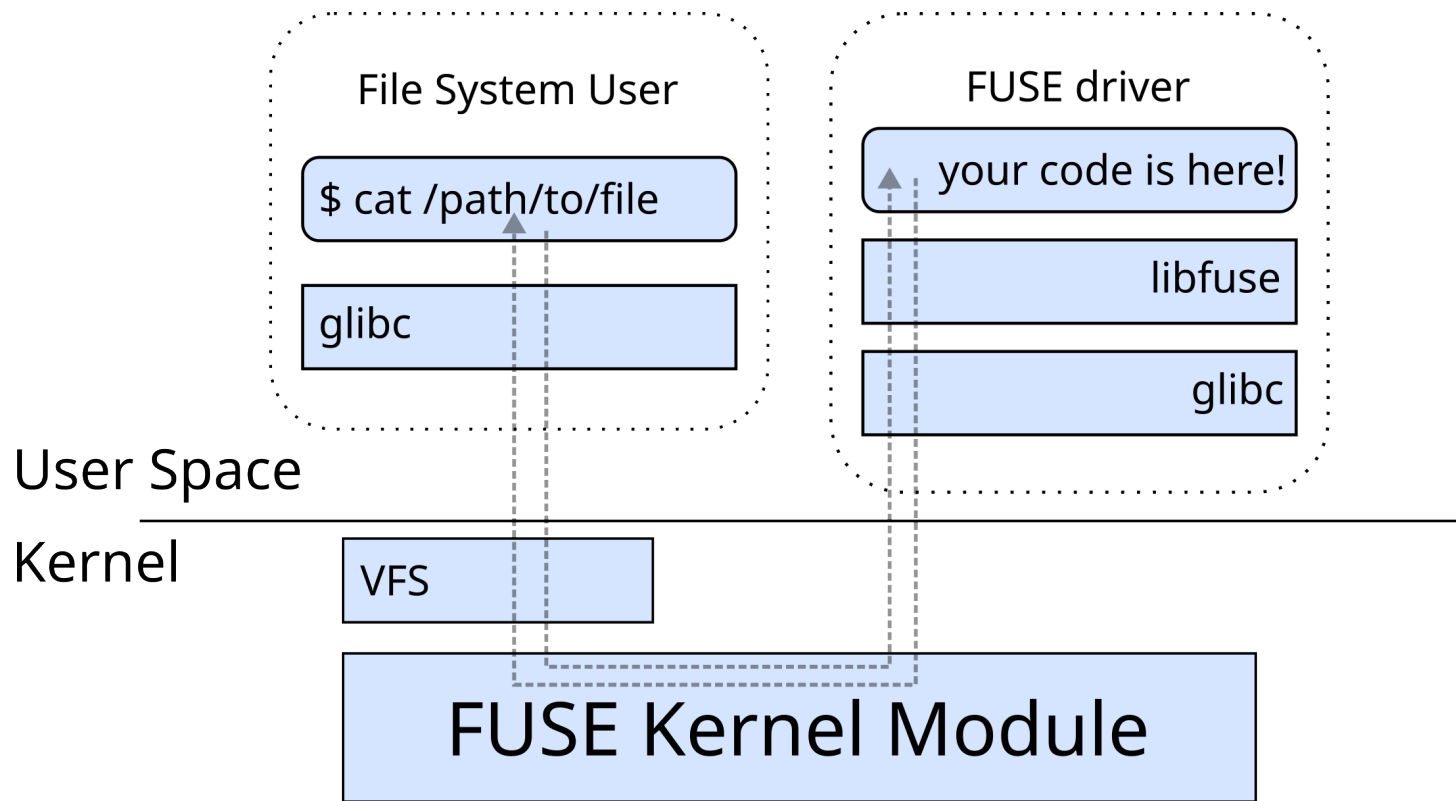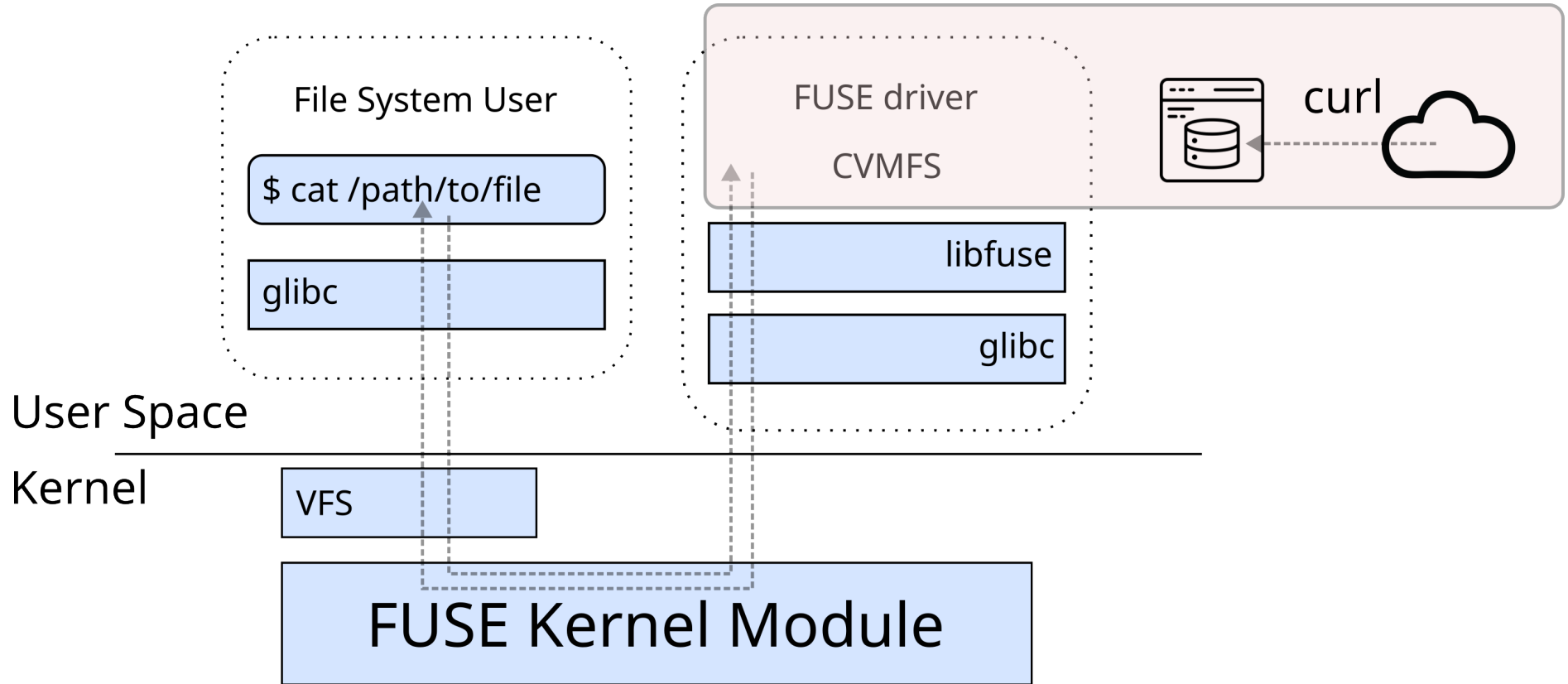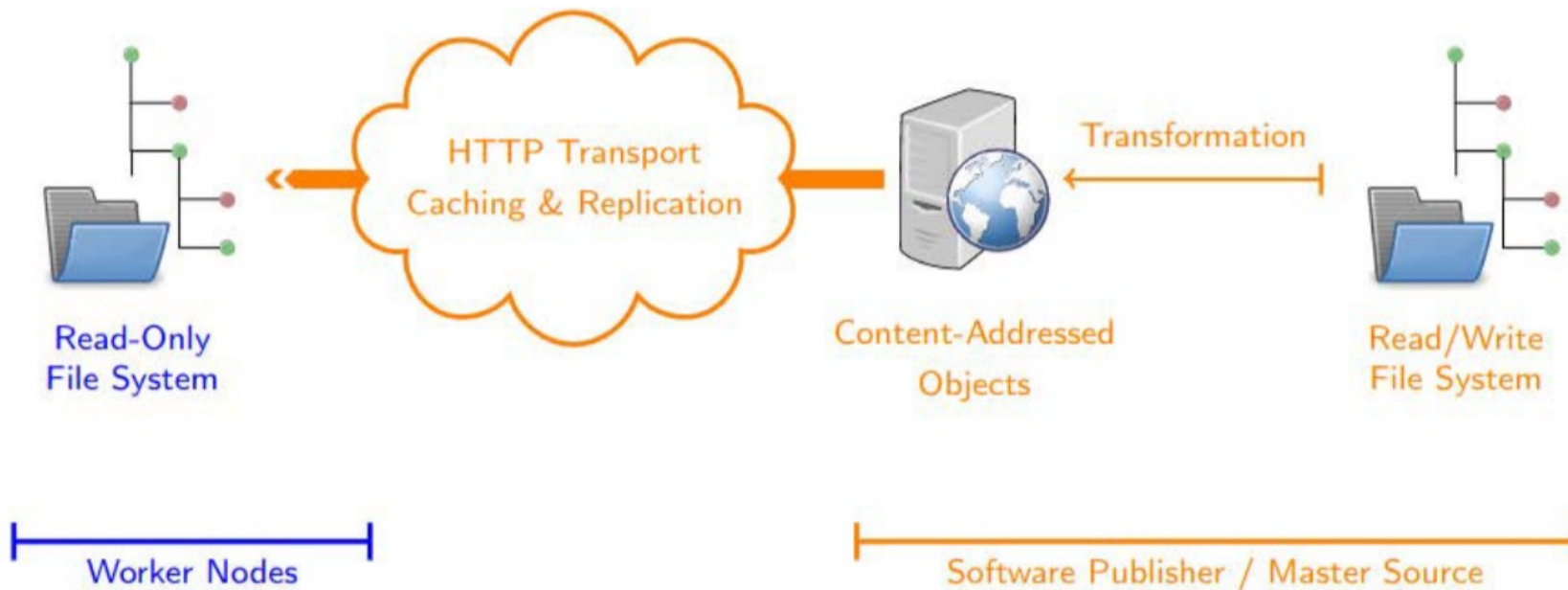
CVMFS is a filesystem in user space

CVMFS is a filesystem in user space

User Space
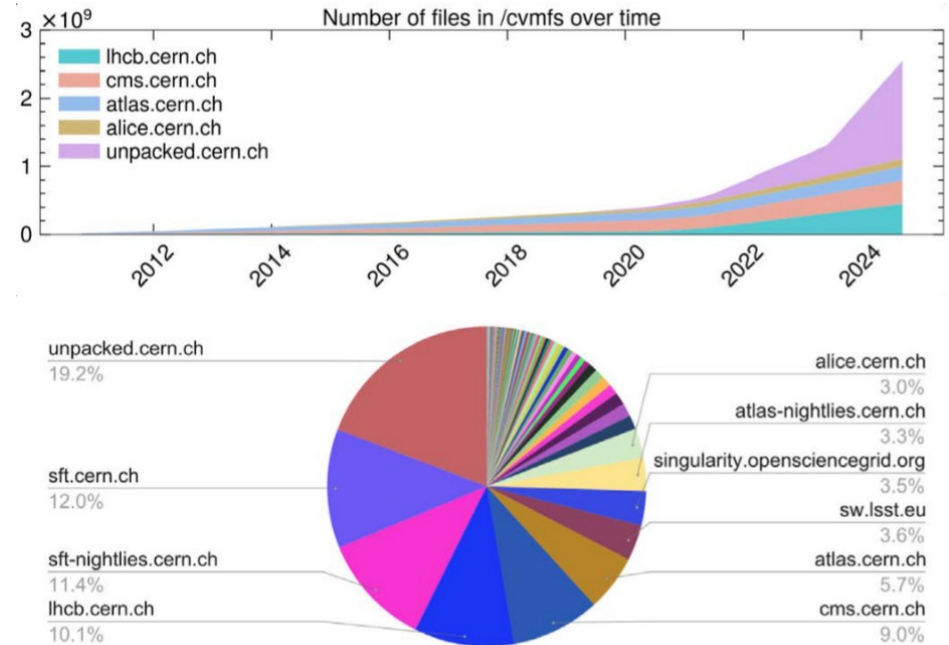
Kernel

File distribution

- Files are stored by content using hash (Merkle) trees

- Allows for data verification and compression

- LRU cache on client side

- Content addressable storage allows **de-duplication** (each hash is stored only once)

HTTP Transport Caching & Replication

Transformation

Read-Only File System

Content-Addressed Objects

Read/Write File System

Worker Nodes

Software Publisher / Master Source
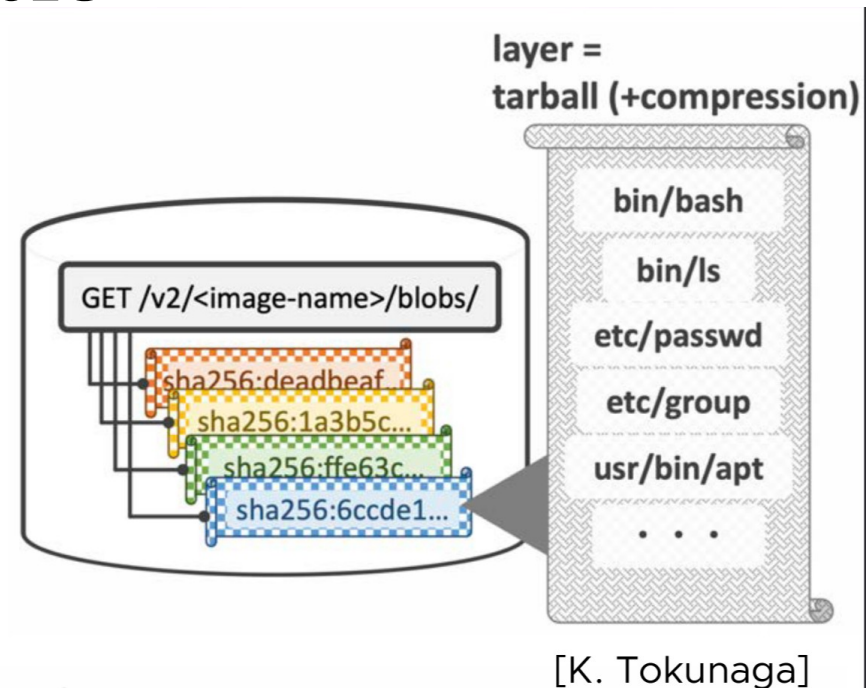
CVMFS at CERN with numbers

- > 33B files

- > 2PB of data

- > 8k container images

- ~ 290 Repositories

A well established tool for
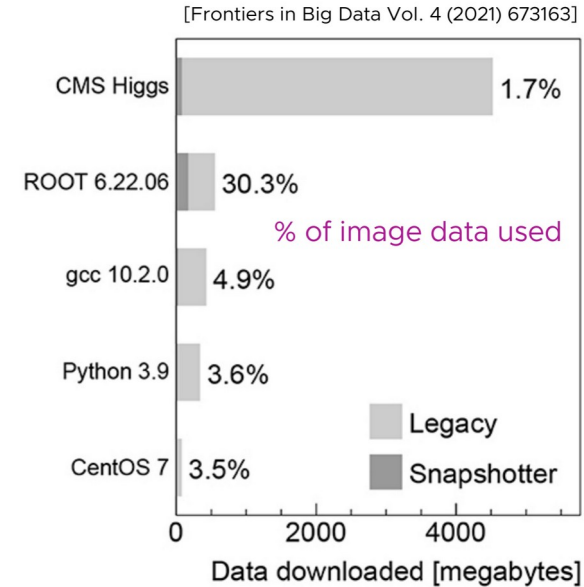software distribution
(and preservation)!

Distribution of Container Images

Images are **unseekable**
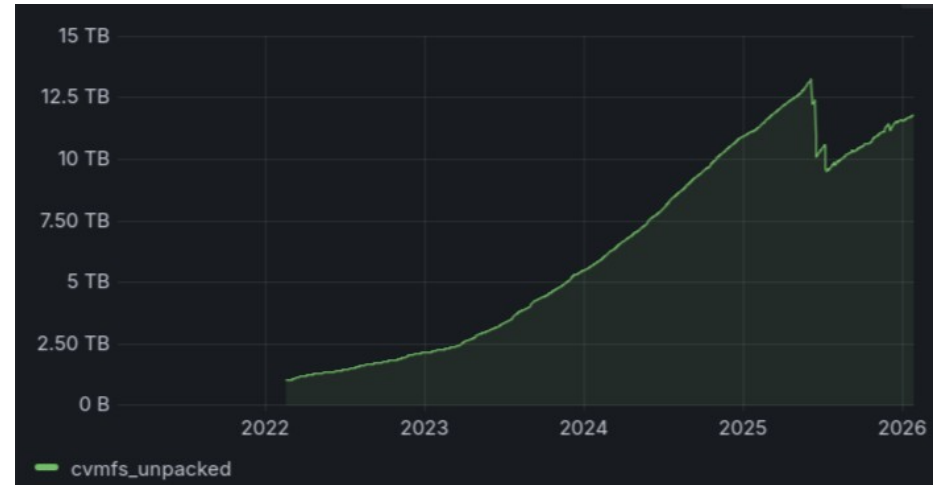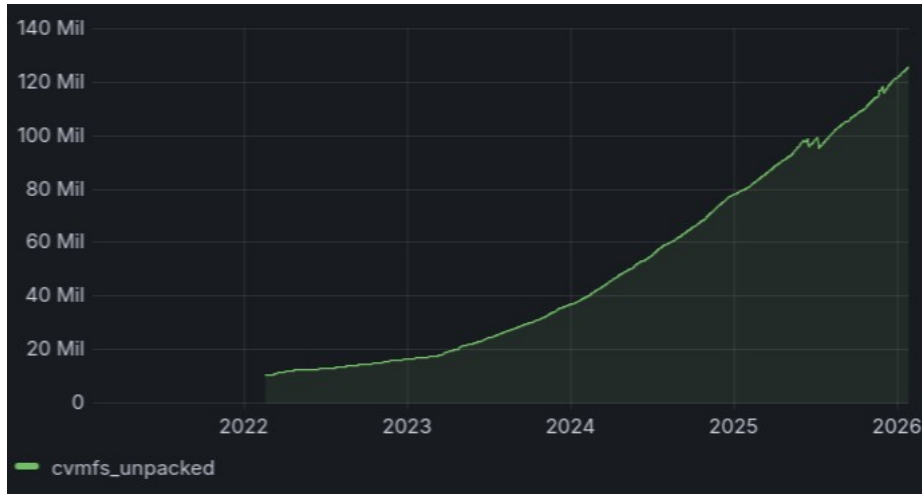and **unindexed** file
archives.

layer =
tarball (+compression)

GET /v2/<image-name>/blobs/

sha256:deadbeaf...
sha256:1a3b5c...
sha256:ffe63c...
sha256:6ccde1...

bin/bash
bin/ls
etc/passwd
etc/group
usr/bin/apt
. . .

[K. Tokunaga]

Lazy pulling

- pull/download the essential parts of the image on demand.

- The container can start almost immediately (possibly slowlier execution)

CMS Higgs — 1.7%

ROOT 6.22.06 — 30.3%

% of image data used

gcc 10.2.0 — 4.9%

Python 3.9 — 3.6%

CentOS 7 — 3.5%

Legacy

Snapshotter

Data downloaded [megabytes]

All fallback to legacy pulling if image not available in required format.

container**d**

10

unpacked.cern.ch

>12 TB of compressed and deduplicated images scattered among >120million files.

(~1.3PB logical data)